

## **Supplementary Materials for Meisner et al.**

### **Supplementary Tables**

Supplementary Table 2: The number of SNPs included in each PRS after removing SNPs in linkage disequilibrium via clumping.

Supplementary Table 3: The estimated association between each mortality risk factor PRS and the risk factor measured at study baseline in women and men.

Supplementary Table 4: The results of the analysis of all-cause mortality and the cPRS fitted in the training data and evaluated in the healthy subset of the test data.

Supplementary Table 5: ICD-10 codes for the top causes of death.

Supplementary Table 6: Methods for identifying prevalent and incident cases of each disease included in the analysis.

Supplementary Table 7: Conversion of self-reported alcohol intake to grams of alcohol per day.

### **Supplementary Figures**

Supplementary Figure 1: The estimated association between each disease PRS and prevalent and incident disease.

Supplementary Figure 2: Cause-specific mortality results, stratified by the presence of disease at study baseline.

Supplementary Figure 3: The estimated association between each mortality risk factor PRS and mortality due to each of the top causes of death among women.

Supplementary Figure 4: The estimated association between each mortality risk factor PRS and mortality due to each of the top causes of death among men.

Supplementary Figure 5: Association of trait-specific PRS with all-cause mortality in models with all 25 PRS.

<b>Trait</b>	<b># SNPs</b>
Alcohol consumption	58
Alzheimer's disease	31
BMI	1,458
Breast cancer	153
CAD	207
Chronic kidney disease	4
Cirrhosis	2
Colorectal cancer	34
COPD	20
DBP	352
eGFR	31
Fasting blood glucose	24
HDL cholesterol	223
Hypertension	7
LDL cholesterol	195
Lung cancer	17
Pancreatic cancer	18
Parkinson's disease	44
Prostate cancer	123
SBP	390
Sleep duration	95
Smoking status	127
Stroke	79
Total cholesterol	240
Triglycerides	138
Type 2 diabetes	175
<i>Total number of unique SNPs</i>	<i>3,941</i>

**Supplementary Table 2:** The number of SNPs included in each PRS after removing SNPs in linkage disequilibrium via clumping. BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; SNP: single nucleotide polymorphism; PRS: polygenic risk score.

<b>Mortality risk factor</b>	<b>Women</b>	<b>Men</b>
Alcohol consumption (grams/day)	0.90 (0.83, 0.96)	1.90 (1.79, 2.02)
BMI (kg/m <sup>2</sup> )	1.46 (1.44, 1.49)	1.26 (1.24, 1.28)
DBP (mm Hg)	1.90 (1.85, 1.95)	1.63 (1.58, 1.68)
eGFR (mL/min/1.73 m <sup>2</sup> )	2.54 (2.47, 2.61)	2.36 (2.28, 2.44)
Blood glucose (mmol/L)	0.065 (0.060, 0.070)	0.077 (0.069, 0.084)
HDL cholesterol (mmol/L)	0.118 (0.117, 0.120)	0.095 (0.094, 0.097)
LDL cholesterol (mmol/L)	0.234 (0.230, 0.238)	0.194 (0.190, 0.198)
SBP (mm Hg)	3.82 (3.74, 3.90)	3.06 (2.98, 3.14)
Sleep duration (hour)	0.092 (0.087, 0.097)	0.082 (0.077, 0.087)
Smoking status (odds ratio for ever smoking)	1.20 (1.19, 1.22)	1.22 (1.21, 1.23)
Total cholesterol (mmol/L)	0.300 (0.295, 0.305)	0.257 (0.251, 0.262)
Triglycerides (mmol/L)	0.187 (0.183, 0.191)	0.269 (0.263, 0.275)

**Supplementary Table 3:** The estimated association between each mortality risk factor PRS and the risk factor measured at study baseline in women and men. Estimates are based on sex-specific linear regression models with robust standard error estimates, with the exception of smoking status, which was modeled using sex-specific logistic regression models. All models included adjustment for age at entry. Estimates are reported per standard deviation of the PRS. BMI: body mass index; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; PRS: polygenic risk score.

	Women	Men
<b>Population (deaths) in test data: N</b>		
Total	29,379 (444)	18,249 (531)
Top 5% of cPRS	1,371 (27)	588 (21)
Middle 20% of cPRS	5,843 (80)	3,680 (107)
Bottom 5% of cPRS	1,647 (21)	1,145 (28)
<b>Summary statistics for test data</b>		
Age at entry (years; mean (SD))	54.4 (7.9)	54.3 (8.2)
Follow-up (years; mean (SD))	8.9 (0.9)	8.8 (1.1)
<b>cPRS: HR (95% CI)</b>		
Per SD of cPRS	1.07 (0.98, 1.18)	1.15 (1.06, 1.26)
Top 5% vs. middle 20% of cPRS	1.46 (0.94, 2.25)	1.28 (0.80, 2.04)
Bottom 5% vs. middle 20% of cPRS	0.89 (0.55, 1.44)	0.78 (0.51, 1.18)
<b>cPRS: years of life lost (95% CI)</b>		
Per SD of cPRS	0.71 (-0.21, 1.63)	1.43 (0.56, 2.31)
Top 5% vs. middle 20% of cPRS	3.75 (-0.61, 8.12)	2.45 (-2.23, 7.13)
Bottom 5% vs. middle 20% of cPRS	-1.16 (-5.97, 3.64)	-2.50 (-6.67, 1.66)

**Supplementary Table 4:** The results of the analysis of all-cause mortality and the cPRS fitted in the training data and evaluated in the healthy subset of the test data. The cPRS were evaluated by fitting sex-specific Cox proportional hazards models of the association between age at death from all causes and the cPRS in the healthy subset of the test data. The healthy subset of the test data was defined as the test data with individuals with any of the diseases included as a top cause of death at baseline (prevalent cases). Both the continuous cPRS and categorical cPRS were modeled. The estimated HRs and CIs were converted to estimated years of life lost. HR: hazard ratio; CI: confidence interval; cPRS: composite PRS; PRS: polygenic risk score; SD: standard deviation.

CDC Definition			Our Definition		
Ranking in US (UK)		Cause	ICD-10 codes	Cause	ICD-10 codes
Women	Men				
1 (2)	1 (2)	Diseases of heart	I00-I09, I11, I13, I20-I51	CAD	I20-I25
2 (1)	2 (1)	Malignant neoplasms	C00-C97	Pancreatic	C25
				Colorectal	C18-C20
				Breast	C50
				Lung	C33-C34
				Prostate	C61
3 (4)	3 (3)	Chronic lower respiratory diseases	J40-J47	Chronic obstructive pulmonary disease	J41-J44
4 (5)	5 (5)	Alzheimer's disease	G30	Alzheimer's disease	G30
5 (3)	4 (4)	Cerebrovascular diseases	I60-I69	Stroke	I60, I61, I63, I64
6 (6)	6 (9)	Diabetes mellitus	E10-E14	Type 2 diabetes	E11
7 (10)	8 (10)	Nephritis, nephrotic syndrome and nephrosis	N00-N07, N17-N19, N25-N27	Chronic kidney disease	N18
8 (11)	10 (11)	Essential hypertension and hypertensive renal disease	I10, I12, I15	Hypertension	I10
9 (7)	7 (6)	Chronic liver disease and cirrhosis	K70, K73-K74	Alcoholic liver cirrhosis	K70.3
10 (8)	9 (7)	Parkinson's disease	G20-G21	Parkinson's disease	G20

**Supplementary Table 5:** ICD-10 codes for the top causes of death identified based on the CDC WONDER database and ICD-10 codes for the corresponding specific cause of death included in the analysis. Ranking in the US based on data for 2017 from CDC WONDER for non-Hispanic whites aged 40 and over; ranking in the UK based on data for 2017 from the Office of National Statistics for individuals aged 40 and over. CDC: Centers for Disease Control; ICD: International Classification of Diseases; WONDER: Wide-ranging ONline Data for Epidemiologic Research.

Cause of death	ICD Codes		Prevalent Definition	Incident Definition
	ICD9	ICD10		
Coronary artery disease	410-414	I20-I25	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported CAD at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Pancreatic cancer	157	C25	Cancer registry: one of the ICD (9/10) codes in the cancer registry with an initial date prior to the date of baseline assessment	Cancer registry: one of the ICD (9/10) codes in the cancer registry with an initial date after date of baseline assessment
Colorectal cancer	153, 154.0, 154.1, 154.8	C18-C20		
Breast cancer	174	C50		
Lung cancer	162	C33-C34		
Prostate cancer	185	C61		
COPD	491, 492, 496	J41-J44	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported COPD, emphysema, or chronic bronchitis at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Alzheimer's disease	331.0	G30 and F00	(a) HES: one of the ICD (9/10) codes in the primary or any secondary position with an initial code date is prior to the date of baseline assessment.	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Stroke	430, 431, 434, 436	I60, I61, I63, I64	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported stroke at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Type 2 diabetes	Defined based on algorithms in Eastwood et al. (1)			
Chronic kidney disease	585	N18	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Hypertension	401	I10	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: (i) self-reported essential hypertension or "any hypertension" but not "gestational hypertension/pre-eclampsia" at baseline or (ii) hypertension medication usage at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death

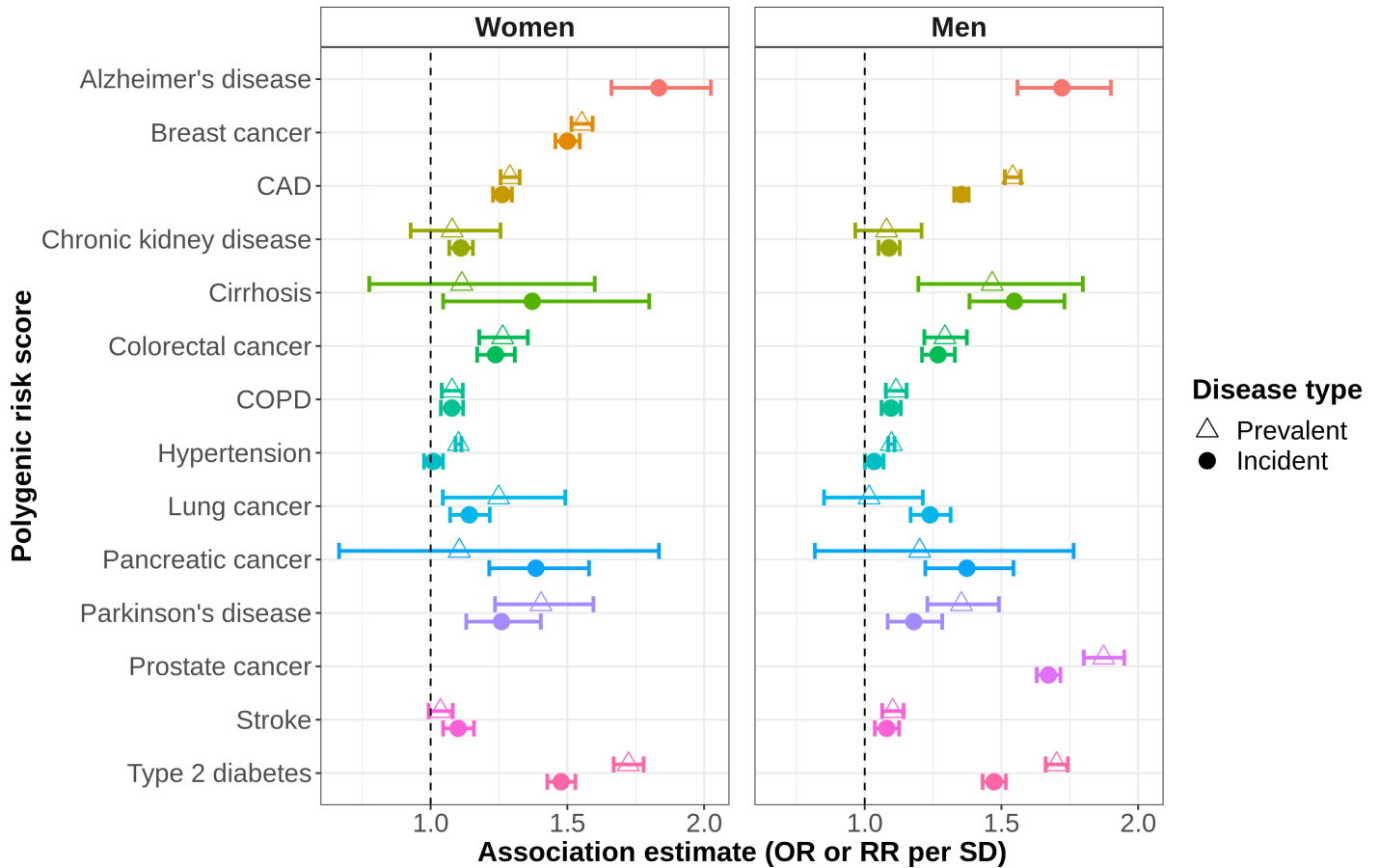
			(c) SBP/DBP measures: systolic blood pressure $\geq 140$ mmHg, or diastolic blood pressure $\geq 90$ mmHg at baseline	
Alcoholic liver cirrhosis	571.2	K70.3	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death
Parkinson's disease	332.0	G20	(a) HES: ICD (9/10) codes in HES in the primary or secondary position with an initial code date prior to the date of baseline assessment (b) Self-report: self-reported Parkinson's disease at baseline	(a) HES: one of the ICD (9/10) codes in HES in the primary or secondary position with an initial code date after the date of baseline assessment (b) Mortality data: one of the ICD10 codes listed as a primary or secondary cause of death

**Supplementary Table 6:** Methods for identifying prevalent and incident cases of each disease included in the analysis. COPD: chronic obstructive pulmonary disease; HES: hospital episode statistics data; ICD: International Classification of Diseases.

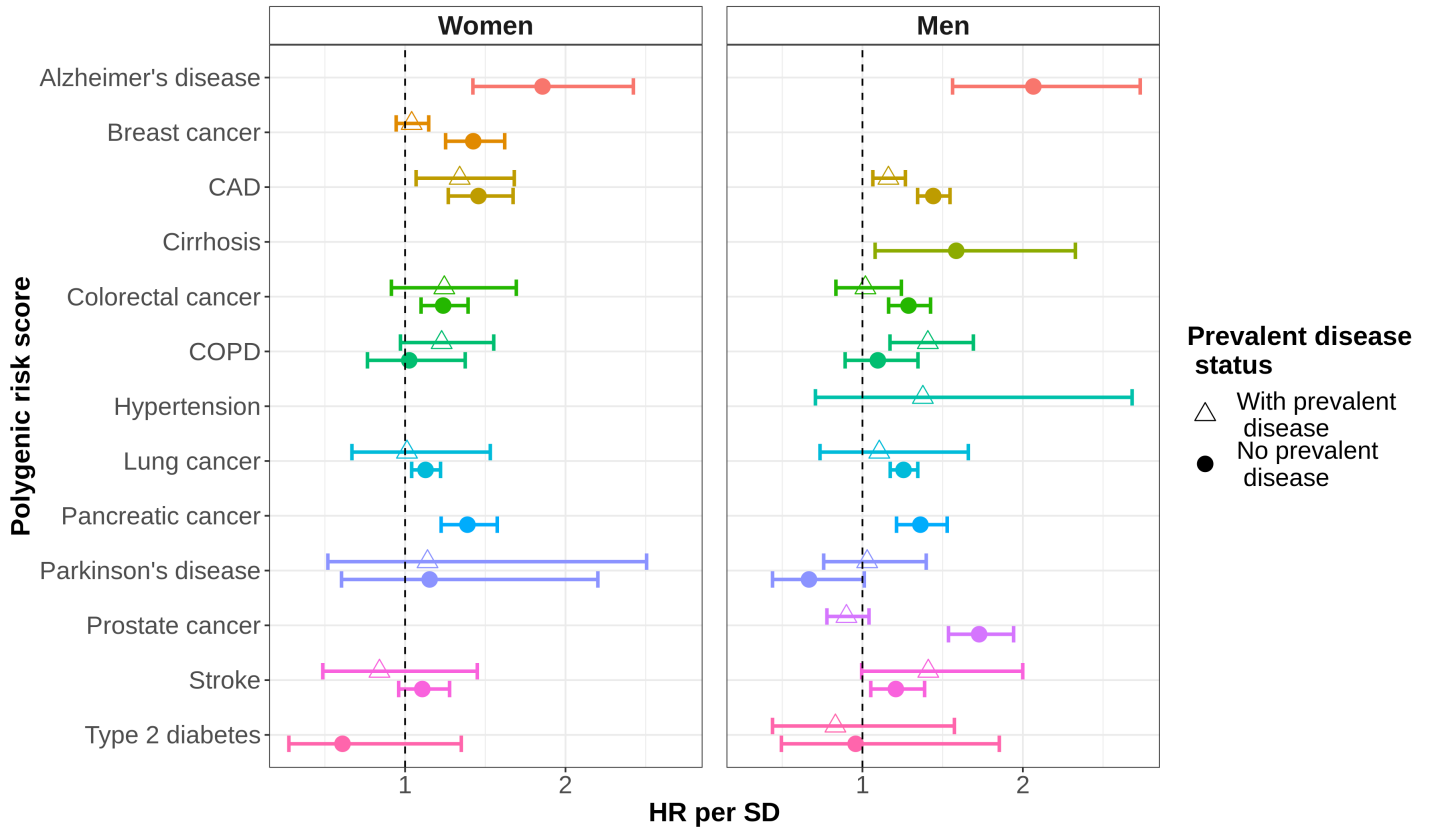
<b>Source</b>	<b>Factor</b>
Red wine intake	1.5
Champagne/white wine	1.5
Beer/cider	2.5
Spirits	1
Fortified wine	1
Other alcoholic drinks	1.5

**Supplementary Table 7:** Conversion of self-reported alcohol intake to grams of alcohol per day. To compute grams of alcohol per day: (1) for each source of alcohol, multiply by the given factor and divide by 7 (if input is weekly intake) or 30 (if input is monthly intake) to get units/day; (2) multiply units/day by 8 to obtain grams/day; (3) sum grams/day intake of each source of alcohol to get total grams of alcohol per day.

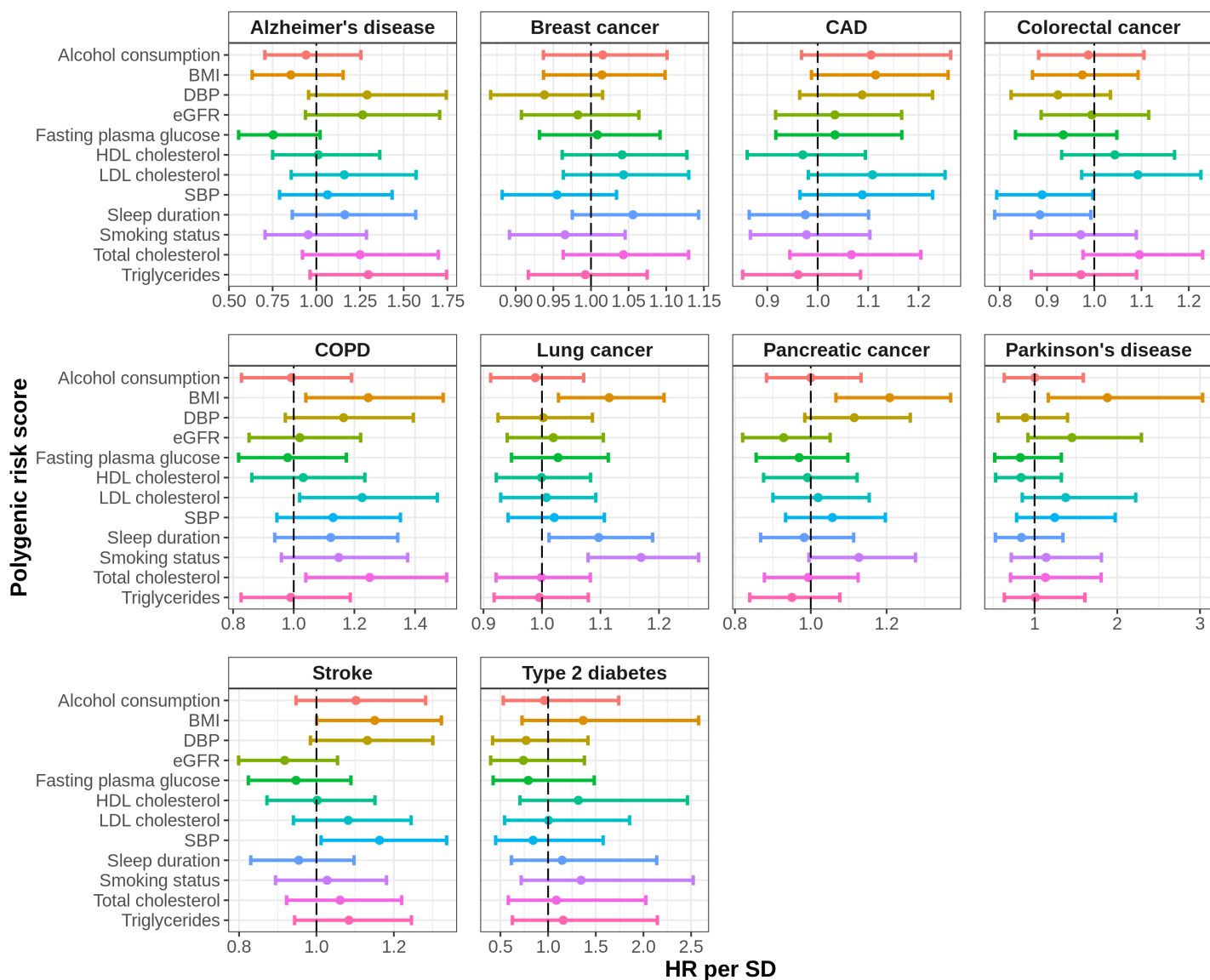




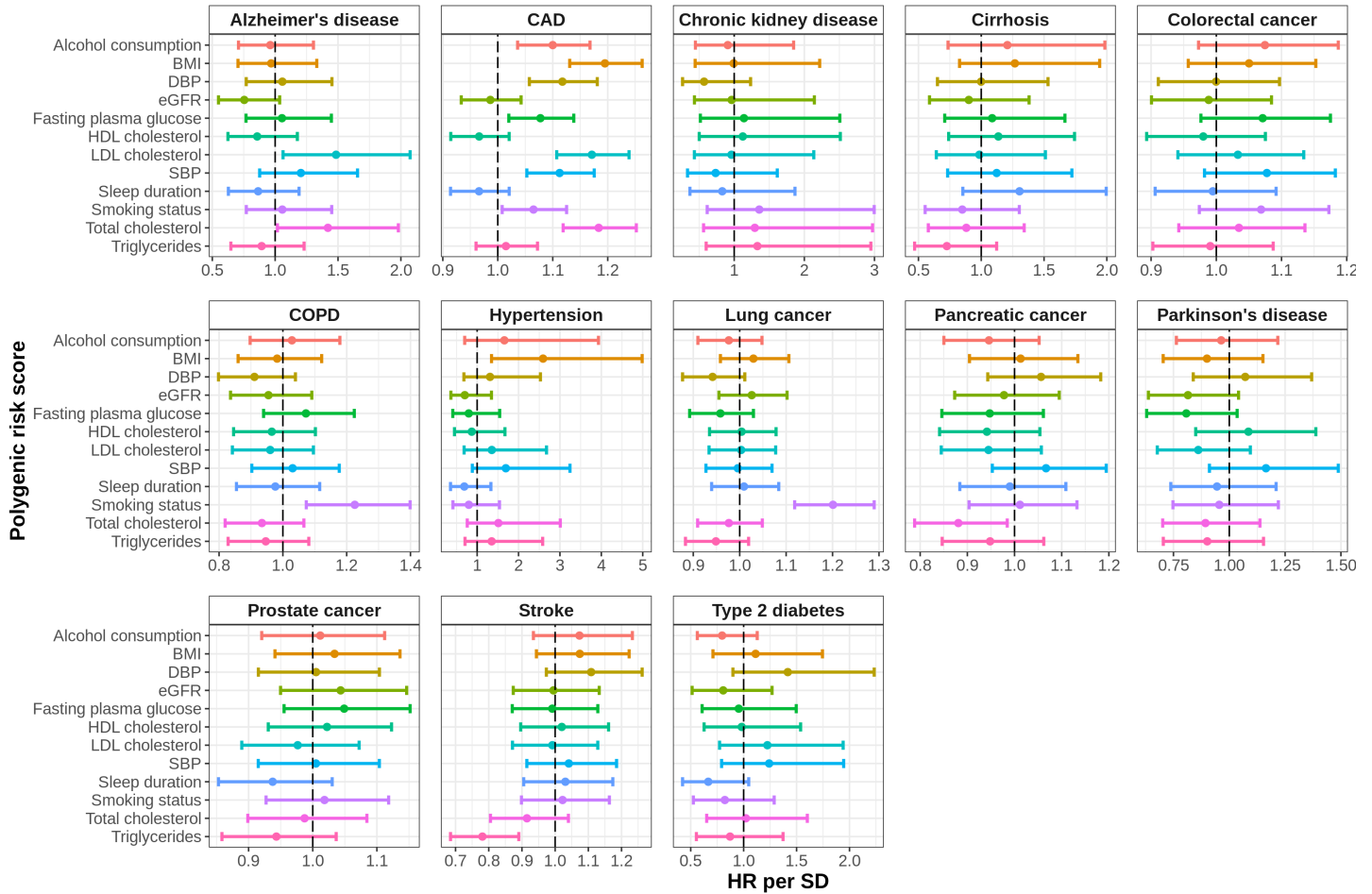
**Supplementary Figure 1:** The estimated association between each disease PRS and prevalent and incident disease in men and women. For prevalent disease (open triangles in the plot), sex-specific logistic regression models were fit in the full cohort. For incident disease (closed circles in the plot), sex-specific modified Poisson regression models with robust standard error estimates were fit to the full cohort, excluding individuals with the disease at baseline (prevalent cases). All models included adjustment for age at entry. The estimates are presented as the estimated OR or RR per standard deviation of the PRS. The horizontal lines indicate 95% confidence intervals. As the number of prevalent cases of Alzheimer's disease was quite low for both men and women, these estimates are not presented. CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; OR: odds ratio; RR: relative risk; SD: standard deviation; PRS: polygenic risk score.



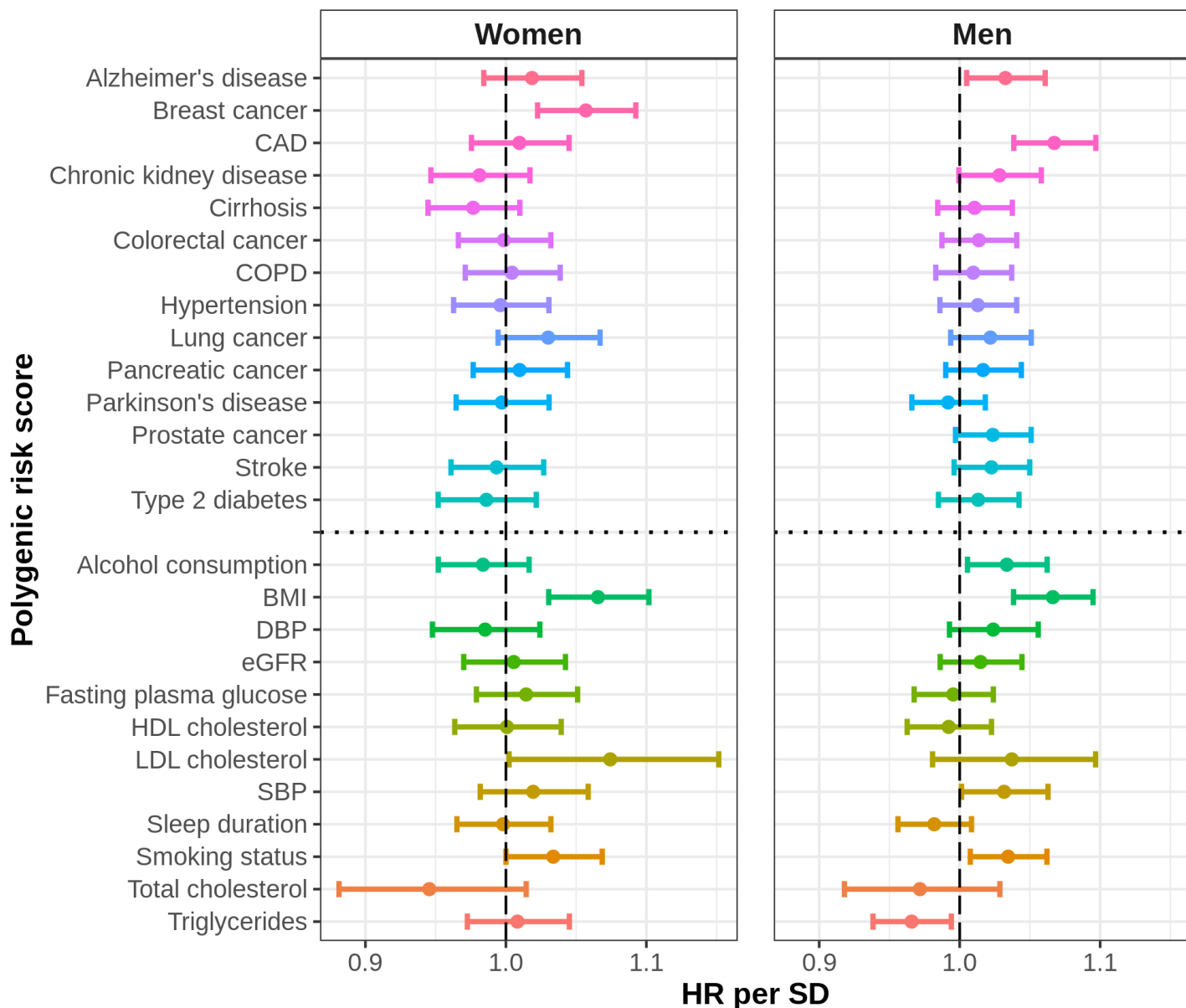
**Supplementary Figure 2:** Cause-specific mortality results in the full cohort in women and men, stratified by the presence of disease at study baseline. For each disease, we evaluated the association between the disease PRS and mortality from the disease based on sex-specific Cox proportional hazards models of age at death in individuals with the disease at baseline (open triangles in the plot) and in individuals without the disease at baseline (closed circles in the plot). Deaths from other causes were treated as censoring events. Some causes did not have enough observations or deaths to yield stable estimates (< 30 observations or < 6 deaths); in these cases, estimates are not provided. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.



**Supplementary Figure 3:** The estimated association between each mortality risk factor PRS and mortality due to each of the top causes of death among women in the full cohort. For each disease, we evaluated the association between each of the risk factor PRS and mortality from the disease based on Cox proportional hazards models of age at death in women. Deaths from other causes were treated as censoring events. Some causes did not have enough deaths to yield stable estimates (< 6 deaths); in these cases, estimates are not provided. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.



**Supplementary Figure 4:** The estimated association between each mortality risk factor PRS and mortality due to each of the top causes of death among men in the full cohort. For each disease, we evaluated the association between each of the risk factor PRS and mortality from the disease based on Cox proportional hazards models of age at death in men. Deaths from other causes were treated as censoring events. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.



**Supplementary Figure 5:** Association of trait-specific PRS with all-cause mortality in the training data based on models with all PRS; these association estimates were used to weight each PRS to form the cPRS. The estimates are based on sex-specific Cox proportional hazards models of age at death with all 25 PRS, fit in the training data. Each PRS was standardized to have unit variance so the estimates correspond to the HR per SD of the PRS. The horizontal lines indicate 95% confidence intervals. BMI: body mass index; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; DBP: diastolic blood pressure; eGFR: estimated glomerular filtration rate; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SBP: systolic blood pressure; HR: hazard ratio; SD: standard deviation; PRS: polygenic risk score.

## References

1. Eastwood S V., Mathur R, Atkinson M, Brophy S, Sudlow C, Flaig R, et al. Algorithms for the capture and adjudication of prevalent and incident diabetes in UK Biobank. PLoS One. 2016 Sep 1;11(9).